

ON THE NORM CONVERGENCE OF NONCONVENTIONAL ERGODIC AVERAGES

Tim Austin

Abstract

We offer a proof of the following nonconventional ergodic theorem:

Theorem. *If $T_i : \mathbb{Z}^r \curvearrowright (X, \Sigma, \mu)$ for $i = 1, 2, \dots, d$ are commuting probability-preserving \mathbb{Z}^r -actions, $(I_N)_{N \geq 1}$ is a Følner sequence of subsets of \mathbb{Z}^r , $(a_N)_{N \geq 1}$ is a base-point sequence in \mathbb{Z}^r and $f_1, f_2, \dots, f_d \in L^\infty(\mu)$ then the nonconventional ergodic averages*

$$\frac{1}{|I_N|} \sum_{n \in I_N + a_N} \prod_{i=1}^d f_i \circ T_i^n$$

converge to some limit in $L^2(\mu)$ that does not depend on the choice of $(a_N)_{N \geq 1}$ or $(I_N)_{N \geq 1}$.

The leading case of this result, with $r = 1$ and the standard sequence of averaging sets, was first proved by Tao in [16], following earlier analyses of various more special cases and related results by Conze and Lesigne [4, 5, 6], Furstenberg and Weiss [9], Zhang [18], Host and Kra [12, 13], Frantzikinakis and Kra [7] and Ziegler [19]. While Tao's proof rests on a conversion to a finitary problem, we invoke only techniques from classical ergodic theory, so giving a new proof of his result.

1 Introduction

The setting for this work is a collection of d commuting measure-preserving actions $T_i : \mathbb{Z}^r \curvearrowright (X, \Sigma, \mu)$, $i = 1, 2, \dots, d$, on a probability space. We present a proof of the following result:

Theorem 1.1 (Convergence of multidimensional nonconventional ergodic averages). *If $T_i : \mathbb{Z}^r \curvearrowright (X, \Sigma, \mu)$ for $i = 1, 2, \dots, d$ are commuting probability-preserving \mathbb{Z}^r -actions, $(I_N)_{N \geq 1}$ is a Følner sequence of subsets of \mathbb{Z}^r , $(a_N)_{N \geq 1}$ is a base-point sequence in \mathbb{Z}^r and $f_1, f_2, \dots, f_d \in L^\infty(\mu)$ then the nonconventional ergodic averages*

$$\frac{1}{|I_N|} \sum_{n \in I_N + a_N} \prod_{i=1}^d f_i \circ T_i^n$$

converge to some limit in $L^2(\mu)$ that does not depend on the choice of $(a_N)_{N \geq 1}$ or $(I_N)_{N \geq 1}$.

The case of this result with $r = 1$ and the standard sequence of averaging sets $I_N + a_N := \{1, 2, \dots, N\}$ was first proved by Tao in [16]. Tao proceeds by first demonstrating the equivalence of this result with a finitary assertion about the behaviour of the restriction of our functions to large finite pieces of individual orbits. This, in turn, is easily seen to be equivalent to a purely finitary result about the behaviour of certain sequences of averages of 1-bounded functions on $(\mathbb{Z}/N\mathbb{Z})^d$ for very large N , and the bulk of Tao's work then goes into proving this last result. Interestingly, Towsner has shown in [17] how the asymptotic behaviour of these purely finitary averages can be re-interpreted back into an ergodic-theoretic assertion by building a suitable 'proxy' probability-preserving system from these averages themselves, using constructions from nonstandard analysis. Tao's method of analysis can be extended to the case of individual actions T_i of a higher-rank r and an arbitrary Følner sequence in \mathbb{Z}^r , but with the base-point shifts a_N all zero, quite straightforwardly, but seems to require more work in order to be extended to a proof for the above base-point-uniform version.

In this paper we shall give a different proof of Theorem 1.1 that uses only more traditional infinitary techniques from ergodic theory. Our method is not affected by shifting the base points of our averages. In particular, we recover a new proof of the base-point-fixed case.

The further special case of Theorem 1.1 in which $r = 1$ and $T_i = T^{a_i}$ for some fixed invertible probability-preserving transformation T and sequence of integers a_1, a_2, \dots, a_d has been the subject of considerable recent attention, with complete proofs of this case appearing in works of Host and Kra [13] and of Ziegler [20]. These, in turn, build on techniques developed in previous papers for this or other special cases of the theorem by Conze and Lesigne [4, 5, 6], Zhang [18] and Host and Kra [12], and also on the analysis by Furstenberg and Weiss in [9] of averages

of the form $\frac{1}{N} \sum_{n=1}^N f \circ T^n \cdot g \circ T^{n^2}$ (which, we stress, do *not* constitute a special case of Theorem 1.1 in view of the nonlinearity in the second exponent).

It is this last paper that first formally introduces the important notion of ‘characteristic factors’ for a system of averages of products: in our general setting, these comprise a tuple $(\Xi_1, \Xi_2, \dots, \Xi_d)$ of T -invariant σ -subalgebras of Σ such that, firstly,

$$\frac{1}{|I_N|} \sum_{n \in I_N + a_N} \prod_{i=1}^d f_i \circ T_i^n - \frac{1}{|I_N|} \sum_{n \in I_N + a_N} \prod_{i=1}^d \mathbb{E}_\mu[f_i | \Xi_i] \circ T_i^n \rightarrow 0$$

in $L^2(\mu)$ as $N \rightarrow \infty$ for any $f_1, f_2, \dots, f_d \in L^\infty(\mu)$ and any choice of $(a_N)_{N \geq 1}$ and $(I_N)_{N \geq 1}$, so that convergence in general will follow if it can be established when each f_i is Ξ_i -measurable; and secondly such that these factors have a more precisely-describable structure than the overall original system, so that the asymptotic behaviour of the right-hand averages above can be analyzed explicitly.

This proof-scheme has not yet been successfully carried out in the general setting of the present paper. The analyses of powers of a single transformation by Host and Kra and by Ziegler both rely on achieving a very precise classification of all possible characteristic factors in the form of ‘nilsystems’, within which setting a bespoke analysis of the convergence of the relevant ergodic averages has been carried out separately by Leibman [14]. In addition, Frantzikinakis and Kra have shown in [7] that nilsystems re-appear in this rôle in the case of a more general collection of invertible single transformations T_i under the assumption that each T_i and each difference $T_i T_j^{-1}$ for $i \neq j$ is ergodic, and they deduce the restriction of Theorem 1.1 to this case also. However, without this extra ergodicity hypothesis simple examples show that any tuple of characteristic factors for our system must be much more complicated, and no good description of such a tuple is known.

We note in passing that in the course of their analysis in [13] of the case of powers of a single transformation, Host and Kra also introduce the following ‘cuboidal’ averages associated to a single action $S : \mathbb{Z}^r \curvearrowright (X, \Sigma, \mu)$:

$$\frac{1}{|I_N|} \sum_{n \in a_N + I_N} \prod_{\eta_1, \eta_2, \dots, \eta_r \in \{0, 1\}} f_\eta \circ S^{\eta_1 n_1 + \eta_2 n_2 + \dots + \eta_r n_r}.$$

Using their structural results they are able to prove convergence of these averages also. This result amounts to a different instance of our Theorem 1.1, involving 2^r commuting \mathbb{Z}^r -actions, by defining $T_\eta^n := S^{\eta_1 n_1 + \eta_2 n_2 + \dots + \eta_r n_r}$.

In this paper we shall use the possibility of projecting our input functions f_i onto special factors only in a rather softer way than in the works above. Noting that the case $d = 1$ of Theorem 1.1 amounts simply to the von Neumann mean ergodic theorem, we shall show that, if $d \geq 2$, and under the assumption that Theorem 1.1 holds for collections of $d - 1$ commuting \mathbb{Z}^r -actions, then from an arbitrary \mathbb{Z}^d -system (X, Σ, μ, T) we can always construct an extension $(\tilde{X}, \tilde{\Sigma}, \tilde{\mu}, \tilde{T})$ and then a factor $\tilde{\Xi}$ of that extension such that, interpreting our nonconventional averages as living inside the larger system \tilde{X} , we may replace the first function f_1 with its projection $E_\mu[f_1 | \tilde{\Xi}]$ in the evaluation of these averages, and this projection is then of such a form that our nonconventional averages can be immediately approximated by nonconventional averages involving only $d - 1$ actions. From this point a proof of Theorem 1.1 follows quickly by induction on d .

It is interesting to note that this overall scheme of building an extension to a system with a certain additional property and then showing that this enables us to project just one of the functions contributing to our nonconventional averages onto a special factor of that extension is the same as that followed by Furstenberg and Weiss in [9]. However, the demands they make on their extension and the ways in which they then exploit it are very different from ours, and at the level of finer detail there seems to be no overlap between the proofs.

In fact, the resulting proof of convergence is much more direct than those previously discovered for the case of powers of a single transformation (in addition to avoiding Tao's conversion to a finitary problem). This is possibly not so surprising: the construction we use to build our extended system $(\tilde{X}, \tilde{\Sigma}, \tilde{\mu}, \tilde{T})$ will typically not respect any additional algebraic structure among the transformations T_i . Even if these are powers of a single transformation, in general the \tilde{T}_i will not be, and thus as far as our proof is concerned this extra assumption lends us no advantage. This is symptomatic of an important price that we pay in following our shorter proof: unlike Host and Kra and Ziegler, we obtain essentially no additional information about the final form that our nonconventional averages take. We suspect that substantial new machinery will be needed in order to describe these limits with any precision.

Finally, let us take this opportunity to stress that the substructures of a system (X, Σ, μ, T) that are responsible for this complexity in the analysis of nonconventional analysis, although complicated and difficult to describe, are in a sense very rare. This heuristic is made precise in the following observation: if the action T is chosen generically (using the coarse topology on the collection of probability-

preserving actions on a fixed Lebesgue space (X, Σ, μ) , say), then classical arguments (see, for example, Chapter 8 of Nadkarni [15]) show that generically every T^γ is individually weakly mixing, and in this case not only can our averages be shown to converge using a rather shorter argument (due to Bergelson in [1]), but they converge simply to the product of the separate averages, $\prod_{i=1}^d \int_X f_i d\mu$. We should like to propose a view of the present paper as a contribution to understanding those rare, specially structured ways in which the averages associated to our system can deviate from this ‘purely random’ behaviour.

Acknowledgements My thanks go to Vitaly Bergelson, John Griesmer, Bernard Host, Bryna Kra, Terence Tao and Tamar Ziegler for several helpful discussions and communications and to David Fremlin and an anonymous referee for several constructive suggestions for improvement.

2 Some preliminary definitions and results

Our interest in this paper is with a probability-preserving system $T : \mathbb{Z}^{rd} \curvearrowright (X, \Sigma, \mu)$, for which we will always assume that the underlying measurable space is standard Borel. Inside \mathbb{Z}^{rd} we distinguish the subgroups $\Gamma_1 := \mathbb{Z}^r \times \{0\}^{r(d-1)}$, $\Gamma_2 := \{0\}^r \times \mathbb{Z}^r \times \{0\}^{r(d-2)}$, \dots and $\Gamma_d := \{0\}^{r(d-1)} \times \mathbb{Z}^r$. Each of these is canonically isomorphic to \mathbb{Z}^r when written as a Cartesian product, as here, and we write $\alpha_i : \mathbb{Z}^r \xrightarrow{\cong} \Gamma_i$ for these isomorphisms. We identify the restrictions $T|_{\Gamma_1}, T|_{\Gamma_2}, \dots, T|_{\Gamma_d}$ with the individual \mathbb{Z}^r actions $T_i^{\alpha_i(\cdot)}$, and denote them by T_1, T_2, \dots, T_d respectively. Note that, in this setting of group actions, all of our transformations are implicitly invertible; routine arguments easily recover versions of Theorem 1.1 suitable for collections of commuting non-invertible transformations. We shall sometimes denote a probability-preserving system alternatively by (X, Σ, μ, T) .

We shall also handle several μ -complete T -invariant σ -subalgebras of Σ . As is standard in ergodic theory we shall use the term **factor** either for such a σ -subalgebra or for a probability-preserving intertwining map $\phi : (X, \Sigma, \mu, T) \rightarrow (Y, \Xi, \nu, S)$; to any such ϕ we can associate the invariant σ -subalgebra given by the μ -completion of $\phi^{-1}[\Xi]$ inside Σ . Henceforth we shall abusively write $\phi^{-1}[\Xi]$ for this completed σ -algebra.

In particular, within our system we can identify the **invariant factor** comprising

all $A \in \Sigma$ such that $\mu(T(A) \triangle A) = 0$. This naturally inherits a \mathbb{Z}^{rd} -action from the original system. We shall denote it by Σ^T . More generally, if Γ is a subgroup of \mathbb{Z}^{rd} , we can identify the factor left invariant by $\{T^\gamma : \gamma \in \Gamma\}$: extending the above notation, we shall call this the $T|_\Gamma$ -**isotropy factor** and write it $\Sigma^{T|_\Gamma}$. We shall frequently refer to this factor in case Γ is the subgroup $\{\alpha_i(n) - \alpha_j(n) : n \in \mathbb{Z}^r\}$ for some $i \neq j$, in which case we write $\Sigma^{T_i=T_j}$ in place of $\Sigma^{T|_{\text{im}(\alpha_i - \alpha_j)}}$. It will be centrally important throughout this paper that if Γ is Abelian then the isotropy factors $\Sigma^{T|_\Gamma}$ are \mathbb{Z}^d -invariant for all $\Gamma \leq \mathbb{Z}^d$; for more general group actions this invariance holds only if Γ is a normal subgroup.

We will assume familiarity with the product measurable space $(X_1 \times X_2 \times \cdots \times X_d, \Sigma_1 \otimes \Sigma_2 \otimes \cdots \otimes \Sigma_d)$ associated to a family of measurable spaces (X_i, Σ_i) , $i = 1, 2, \dots, d$. Given measurable maps $\psi_i : X_i \rightarrow Y_i$ between such spaces we shall write $\psi_1 \times \psi_2 \times \cdots \times \psi_d$ for their coordinate-wise product:

$$\psi_1 \times \psi_2 \times \cdots \times \psi_d(x_1, x_2, \dots, x_d) := (\psi_1(x_1), \psi_2(x_2), \dots, \psi_d(x_d)).$$

More generally, if $T_i : \mathbb{Z}^r \curvearrowright (X_i, \Sigma_i)$ is an action for $i = 1, 2, \dots, d$ then we shall write $T_1 \times T_2 \times \cdots \times T_d$ for the action $\mathbb{Z}^r \curvearrowright (X_1 \times X_2 \times \cdots \times X_d, \Sigma_1 \otimes \Sigma_2 \otimes \cdots \otimes \Sigma_d)$ given by

$$(T_1 \times T_2 \times \cdots \times T_d)^n := T_1^n \times T_2^n \times \cdots \times T_d^n.$$

If all the X_i are equal to X , all the Y_i to Y and all the ψ_i to ψ then we shall abbreviate $\psi \times \psi \times \cdots \times \psi$ to $\psi^{\times d}$, and similarly for actions.

The construction that we later use for our proof of Theorem 1.1 will also require the standard notion of an inverse limit of probability-preserving systems; these are treated, for example, in Examples 6.3 and Proposition 6.4 of Glasner [10]. In addition to the results contained there, we need the following simple lemmas.

Lemma 2.1 (Isotropy factors respect inverse limits). *Suppose that*

$$(X, \Sigma, \mu, T) = \lim_{m \leftarrow} (X^{(m)}, \Sigma^{(m)}, \mu^{(m)}, T^{(m)})$$

is an inverse limit of an increasing sequence of \mathbb{Z}^{rd} -systems with connecting maps $\theta_{(m)}^{(m')} : X^{(m')} \rightarrow X^{(m)}$ for $m' \geq m$ and overall projections $\theta_{(m)} : X \rightarrow X^{(m)}$, and that $\Gamma \leq \mathbb{Z}^{rd}$. Then

$$\Sigma^{T|_\Gamma} = \bigvee_{m \geq 1} \theta_{(m)}^{-1}[(\Sigma^{(m)})^{T^{(m)}|_\Gamma}].$$

Proof It is clear that $\Sigma^{T|_\Gamma} \supseteq \theta_{(m)}^{-1}[(\Sigma^{(m)})^{T^{(m)}|_\Gamma}]$ for every $m \geq 1$, and therefore that $\Sigma^{T|_\Gamma} \supseteq \bigvee_{m \geq 1} \theta_{(m)}^{-1}[(\Sigma^{(m)})^{T^{(m)}|_\Gamma}]$; it remains to prove the reverse inclusion. Thus, suppose that $A \in \Sigma$ is $T|_\Gamma$ -invariant. Then, by the construction of the inverse limit, for any $\varepsilon > 0$ we can pick some $m_\varepsilon \geq 1$ and some $A_\varepsilon \in \theta_{(m_\varepsilon)}^{-1}[\Sigma^{(m_\varepsilon)}]$ with $\mu(A \triangle A_\varepsilon) < \varepsilon$. This last inequality is equivalent to $\|1_A - 1_{A_\varepsilon}\|_1 < \varepsilon$. Since A is $T|_\Gamma$ -invariant it follows that $\|1_A - 1_{A_\varepsilon} \circ T^\gamma\|_1 < \varepsilon$ for every $\gamma \in \Gamma$; hence, letting f be the ergodic average of 1_{A_ε} under the action of $T|_\Gamma$, we deduce that $f \in L^\infty(\mu|_{\theta_{(m)}^{-1}[\Sigma^{(m)}]})$, f is $T|_\Gamma$ -invariant and $\|1_A - f\|_1 < \varepsilon$. Now taking a level-set decomposition of f yields $T|_\Gamma$ -invariant sets in $\theta_{(m)}^{-1}[\Sigma^{(m)}]$ that approximate A to within ε . Since ε was arbitrary this shows that A lies in $\bigvee_{m \geq 1} \theta_{(m)}^{-1}[(\Sigma^{(m)})^{T^{(m)}|_\Gamma}]$, as required. \square

Lemma 2.2 (Joins respect inverse limits). *Suppose that (X, Σ, μ) is a probability space and that for each $i = 1, 2, \dots, k$ we have a tower of σ -subalgebras $\Xi_i^{(0)} \subseteq \Xi_i^{(1)} \subseteq \dots \subseteq \Sigma$. Then*

$$\bigvee_{m \geq 1} (\Xi_1^{(m)} \vee \Xi_2^{(m)} \vee \dots \vee \Xi_k^{(m)}) = \left(\bigvee_{m \geq 1} \Xi_1^{(m)} \right) \vee \left(\bigvee_{m \geq 1} \Xi_2^{(m)} \right) \vee \dots \vee \left(\bigvee_{m \geq 1} \Xi_k^{(m)} \right).$$

Proof For every $m \geq 1$ we have

$$\begin{aligned} \Xi_1^{(m)} \vee \Xi_2^{(m)} \vee \dots \vee \Xi_k^{(m)} &\subseteq \left(\bigvee_{m \geq 1} \Xi_1^{(m)} \right) \vee \left(\bigvee_{m \geq 1} \Xi_2^{(m)} \right) \vee \dots \vee \left(\bigvee_{m \geq 1} \Xi_k^{(m)} \right) \\ &\subseteq \bigvee_{m \geq 1} (\Xi_1^{(m)} \vee \Xi_2^{(m)} \vee \dots \vee \Xi_k^{(m)}) \end{aligned}$$

and so taking the limit of the left-hand side above gives the result. \square

3 The Furstenberg self-joining

Central to many of the older ergodic-theoretic analyses of special cases of Theorem 1.1 is a certain multiple self-joining of the input \mathbb{Z}^d -system (X, Σ, μ, T) . Given such a system and also a Følner sequence $(I_N)_{N \geq 1}$ and a base-point sequence $(a_N)_{N \geq 1}$ we can consider the averages

$$\frac{1}{|I_N|} \sum_{n \in a_N + I_N} \int_X \prod_{i=1}^d f_i \circ T_i^n d\mu = \frac{1}{|I_N|} \sum_{n \in a_N + I_N} \int_X f_1 \cdot \prod_{i=2}^d f_i \circ (T_i T_1^{-1})^n d\mu,$$

and now in view of the right-hand expression above, if we know only the rank- $(d-1)$ case of Theorem 1.1 then we can deduce that these averages converge, and it is routine to show (using the standard Borel nature of (X, Σ)) that the resulting limit values define a probability measure μ^{*d} on the product measurable space $(X^d, \Sigma^{\otimes d})$ by the condition that

$$\mu^{*d}(A_1 \times A_2 \times \dots \times A_d) := \lim_{N \rightarrow \infty} \frac{1}{|I_N|} \sum_{n \in a_N + I_N} \int_X \prod_{i=1}^d 1_{A_i} \circ T_i^n d\mu,$$

where we know that this is independent of the choice of $(a_N)_{N \geq 1}$ and $(I_N)_{N \geq 1}$. It is now also clear that this measure μ^{*d} is invariant under the \mathbb{Z}^r -actions $S_i := T_i^{\times d}$ for $i = 1, 2, \dots, d$ and also under $S_{d+1} := T_1 \times T_2 \times \dots \times T_d$. We refer to $(X^d, \Sigma^{\otimes d}, \mu^{*d})$ as the **Furstenberg self-joining** of the space (X, Σ, μ) associated to the action T , in light of its historical genesis in Furstenberg's work on the ergodic theoretic approach to Szemerédi's Theorem ([8]); note, in particular, that the one-dimensional marginals of μ^{*d} on (X, Σ) all coincide with μ . Given this self-joining, we shall write $\pi_1, \pi_2, \dots, \pi_d$ for the projection maps onto the d copies of (X, Σ, μ) that are its coordinate factors.

In the sequel we will need to work simultaneously with the Furstenberg self-joinings of a system (X, Σ, μ, T) and an extension $\psi : (\tilde{X}, \tilde{\Sigma}, \tilde{\mu}, \tilde{T}) \rightarrow (X, \Sigma, \mu, T)$ of that system, in which case we can compute easily that the map $\psi^{\times d}$ identifies $(\tilde{X}^d, \tilde{\Sigma}^{\otimes d}, \tilde{\mu}^{*d})$ as an extension of $(X^d, \Sigma^{\otimes d}, \mu^{*d})$, and we shall write $\tilde{\pi}_1, \tilde{\pi}_2, \dots, \tilde{\pi}_d$ for the coordinate-projections of this larger self-joining.

4 The proof of nonconventional average convergence

We prove Theorem 1.1 by induction on d . As remarked above, the case $d = 1$ is simply the von Neumann mean ergodic theorem, so let us suppose that $d \geq 2$ and that the result is known to be true for all systems of at most $d-1$ commuting \mathbb{Z}^r -actions.

4.1 Characteristic factors and pleasant systems

As indicated in the introduction, we shall use a rather simple instance of the notion of 'characteristic factors':

Definition 4.1 (Characteristic factors). *Given a system $T : \mathbb{Z}^{rd} \curvearrowright (X, \Sigma, \mu)$, a sequence of **characteristic factors** for the nonconventional ergodic averages associated to T_1, T_2, \dots, T_d is a tuple $(\Xi_1, \Xi_2, \dots, \Xi_d)$ of T -invariant σ -subalgebras of Σ such that*

$$\frac{1}{|I_N|} \sum_{n \in a_N + I_N} \prod_{i=1}^d f_i \circ T_i^n - \frac{1}{|I_N|} \sum_{n \in a_N + I_N} \prod_{i=1}^d \mathbb{E}_\mu[f_i | \Xi_i] \circ T_i^n$$

in $L^2(\mu)$ as $N \rightarrow \infty$ for any $f_1, f_2, \dots, f_d \in L^\infty(\mu)$, Følner sequence $(I_N)_{N \geq 1}$ and base-point sequence $(a_N)_{N \geq 1}$.

Many previous results on special cases of Theorem 1.1 have relied on the identification of a tuple of characteristic factors that could then be described quite precisely, in the sense that they can be defined by factor maps of the original system to certain concrete model systems in which a more detailed analysis of nonconventional averages is feasible. Most strikingly, the analysis of Host and Kra in [13] and Ziegler in [20] show that for powers of a single ergodic transformation there is a single minimal characteristic factor (equal to all of the Ξ_i above) that may be identified with a model given by a d -step nilsystem, wherein the convergence of the nonconventional averages and the form of their limits can be analyzed in great detail.

Here we shall not be so ambitious. Various examples show that for a sufficiently complicated system those functions measurable with respect to either Σ^{T_1} or $\Sigma^{T_i=T_1}$ for some $i = 2, 3, \dots, d$ will behave differently (and, in particular, contribute nontrivially) should they appear as f_1 in our averages, and so we expect any tuple of characteristic factors to have $\Xi_1 \supseteq \Sigma^{T_1} \vee \bigvee_{i=2}^d \Sigma^{T_i=T_1}$. In order to explain our approach, let us first suppose that we are given a system in which we may actually take this to be our first characteristic factor, and may simply take $\Xi_i := \Sigma$ for $i = 2, 3, \dots, d$.

Definition 4.2 (Pleasant system). *We shall term a system (X, Σ, μ, T) **pleasant** if*

$$\left(\Sigma^{T_1} \vee \bigvee_{i=2}^d \Sigma^{T_i=T_1}, \Sigma, \Sigma, \dots, \Sigma \right)$$

is a tuple of characteristic factors.

Remark The idea of conditioning just one of the functions f_i in our averages onto a nontrivial factor already appears in Furstenberg and Weiss [9], in whose terminology such a factor is ‘partially characteristic’. \triangleleft

The main observation of this subsection is that, given convergence of nonconventional averages in general for systems of $d - 1$ actions, we can easily deduce that convergence for pleasant systems of d actions. Let us first record separately an elementary robustness result for nonconventional averages that we shall need shortly.

Lemma 4.3. *For any $f_1, f_2, \dots, f_d \in L^\infty(\mu)$ and $N \geq 1$ we have*

$$\left\| \frac{1}{|I_N|} \sum_{n \in a_N + I_N} \prod_{i=1}^d f_i \circ T_i^n \right\|_2 \leq \|f_1\|_2 \cdot \prod_{i=2}^d \|f_i\|_\infty.$$

Proof This is clear from the termwise estimate

$$\left\| \prod_{i=1}^d f_i \circ T_i^n \right\|_2 \leq \|f_1 \circ T_1^n\|_2 \cdot \prod_{i=2}^d \|f_i \circ T_i^n\|_\infty = \|f_1\|_2 \cdot \prod_{i=2}^d \|f_i\|_\infty.$$

and the triangle inequality. \square

Corollary 4.4. *The nonconventional averages*

$$\frac{1}{|I_N|} \sum_{n \in a_N + I_N} \prod_{i=1}^d f_i \circ T_i^n$$

converge in $L^2(\mu)$ for the d -tuple of functions $f_1, f_2, \dots, f_d \in L^\infty(\mu)$ if the corresponding averages are known to converge for all the d -tuples $f_1^{(m)}, f_2, \dots, f_d$ for some sequence $f_1^{(m)} \in L^\infty(\mu)$ that converges to f_1 in $L^2(\mu)$. \square

Proposition 4.5 (Nonconventional average convergence for pleasant systems). *If $T : \mathbb{Z}^d \curvearrowright (X, \Sigma, \mu)$ is pleasant and Theorem 1.1 is known to hold for all systems of $d - 1$ commuting actions, then its conclusion also holds for (X, Σ, μ, T) .*

Proof Writing $\Xi := \Sigma^{T_1} \vee \bigvee_{i=2}^d \Sigma^{T_i=T_1}$, Definition 4.1 tells us that

$$\frac{1}{|I_N|} \sum_{n \in a_N + I_N} \prod_{i=1}^d f_i \circ T_i^n - \frac{1}{|I_N|} \sum_{n \in a_N + I_N} (\mathbb{E}_\mu[f_1 | \Xi] \circ T_1^n) \cdot \prod_{i=2}^d f_i \circ T_i^n \rightarrow 0$$

for any $f_1, f_2, \dots, f_d \in L^\infty(\mu)$, and so it suffices to prove the desired convergence under the additional assumption that f_1 is Ξ -measurable. However, in this case we know that we can approximate f_1 in $L^2(\mu)$ by finite sums of the form $\sum_{k=1}^K g_{1,k} \cdot g_{2,k} \cdot \dots \cdot g_{d,k}$ where $g_{1,k} \in L^\infty(\mu|_{\Sigma^{T_1}})$ and $g_{i,k} \in L^\infty(\mu|_{\Sigma^{T_1=T_i}})$ for $i = 2, 3, \dots, d$. Hence by linearity and Corollary 4.4 it suffices to prove convergence for the averages obtained when f_1 is replaced by a single such product:

$$\frac{1}{|I_N|} \sum_{n \in a_N + I_N} ((g_1 \cdot g_2 \cdot \dots \cdot g_d) \circ T_1^n) \cdot \prod_{i=2}^d f_i \circ T_i^n;$$

but now the different invariances that we are assuming for each g_i imply that $g_1 \circ T_1^n = g_1$ and $g_i \circ T_1^n = g_i \circ T_i^n$ for $i = 2, 3, \dots, d$ and all $n \in \mathbb{Z}^r$, and so the above is simply equal to

$$g_1 \cdot \frac{1}{|I_N|} \sum_{n \in a_N + I_N} \prod_{i=2}^d (g_i \cdot f_i) \circ T_i^n.$$

This is a product by the fixed bounded function g_1 of a nonconventional ergodic average associated to the $d - 1$ commuting actions T_2, T_3, \dots, T_d , and we already know by inductive hypothesis that these converge in $L^2(\mu)$. This completes the proof. \square

Unsurprisingly, there are well-known examples of systems that are unpleasant: for example, the general d -step nilsystems that emerge in the Host-Kra and Ziegler analyses are such. The simplest example from among these is the following: if R_α is an irrational rotation on $(X, \Sigma, \mu) := (\mathbb{T}, \text{Borel}, \text{Haar})$ and we set $T_1 := R_\alpha$, $T_2 := R_{2\alpha} = T_1^2$, then we can check easily that $\Sigma^{T_1} = \Sigma^{T_2} = \Sigma^{T_1=T_2}$ are all trivial, but on the other hand if $f_2 \in \widehat{\mathbb{T}} \setminus \{1_{\mathbb{T}}\}$ and $f_1 := \overline{f_2}^2$ then f_1 and f_2 are both orthogonal to the trivial factor but give

$$\frac{1}{N} \sum_{n=1}^N f_1(T_1^n(t)) f_2(T_2^n(t)) = \frac{1}{N} \sum_{n=1}^N \overline{f_2(t)}^2 f_2(t) \cdot \overline{f_2(\alpha)}^2 f_2(2\alpha) \equiv \overline{f_2(t)} \neq 0$$

as $N \rightarrow \infty$.

However, it turns out that we can repair this situation by passing to a suitable extension.

Proposition 4.6 (All systems have pleasant extensions). *Any \mathbb{Z}^d -system (X, Σ, μ, T) admits a pleasant extension $\psi : (\tilde{X}, \tilde{\Sigma}, \tilde{\mu}, \tilde{T}) \rightarrow (X, \Sigma, \mu, T)$.*

From this point, Theorem 1.1 follows at once, since it is clear that the theorem holds for any system if it holds for some extension of that system. Proposition 4.6 forms the technical heart of this paper, and we shall prove it in the next subsection.

4.2 Building a pleasant extension

We shall build our pleasant extension using the machinery of Furstenberg self-joinings. By the remarks of Section 3, given the conclusions of Theorem 1.1 for systems of $d-1$ commuting \mathbb{Z}^r -actions and a system $T : \mathbb{Z}^{rd} \curvearrowright (X, \Sigma, \mu)$ we may form the Furstenberg self-joining $(X^d, \Sigma^{\otimes d}, \mu^{*d})$. Our deduction of pleasantness for our constructed extension will rest on the following key estimate.

Lemma 4.7 (The Furstenberg self-joining controls nonconventional averages). *If $f_1 \in L^\infty(\mu)$ is such that*

$$\int_{X^d} f_1 \circ \pi_1 \cdot \left(\prod_{i=2}^d f_i \circ \pi_i \right) \cdot g \, d\mu^{*d} = 0$$

*for every choice of $f_2, f_3, \dots, f_d \in L^\infty(\mu)$ and of another function $g \in L^\infty(\mu^{*d}|_{(\Sigma^{\otimes d})^{S_{d+1}}})$, then also*

$$\frac{1}{|I_N|} \sum_{n \in a_N + I_N} \prod_{i=1}^d f_i \circ T_i^n \rightarrow 0$$

in $L^2(\mu)$ for every choice of $f_2, f_3, \dots, f_d \in L^\infty(\mu)$ and any Følner sequence $(I_N)_{N \geq 1}$ and base-point sequence $(a_N)_{N \geq 1}$.

Remark Versions of this result have appeared repeatedly in previous analyses of more special cases of our main result; consider, for example, Proposition 5.3 of Zhang [18] or Subsection 6.3 of Ziegler [20]. The standard proof applies essentially unchanged in the general setting, and we include the details here largely for completeness. \triangleleft

Proof Suppose that $f_1, f_2, \dots, f_d \in L^\infty(\mu)$ satisfy the assumptions of the theorem. By the classical higher-rank van der Corput Lemma (see, for example, the discussion in Bergelson, McCutcheon and Zhang [3]) applied to the bounded

\mathbb{Z}^r -indexed family $\prod_{i=1}^d f_i \circ T_i^n$ in $L^2(\mu)$ we need only prove that

$$\begin{aligned} & \frac{1}{M^{2r}} \sum_{m_1, m_2 \in \{1, 2, \dots, M\}^r} \frac{1}{|I_N|} \sum_{n \in a_N + I_N} \int_X \prod_{i=1}^d (f_i \circ T_i^{m_1+n} \cdot f_i \circ T_i^{m_2+n}) d\mu \\ &= \frac{1}{M^{2r}} \sum_{m_1, m_2 \in \{1, 2, \dots, M\}^r} \frac{1}{|I_N|} \sum_{n \in a_N + I_N} \int_X \prod_{i=1}^d (f_i \circ T_i^{m_1} \cdot f_i \circ T_i^{m_2}) \circ T_i^n d\mu \rightarrow 0 \end{aligned}$$

as $N \rightarrow \infty$ and then $M \rightarrow \infty$. However, by the definition of the Furstenberg self-joining we know that

$$\begin{aligned} & \frac{1}{|I_N|} \sum_{n \in a_N + I_N} \int_X \prod_{i=1}^d (f_i \circ T_i^{m_1} \cdot f_i \circ T_i^{m_2}) \circ T_i^n d\mu \\ & \rightarrow \int_X \prod_{i=1}^d (f_i \cdot f_i \circ T_i^{m_2-m_1}) \circ \pi_i d\mu^{*d} \end{aligned}$$

as $N \rightarrow \infty$. Now, when we take the averages of these limiting values over m_1 and $m_2 \in \{1, 2, \dots, M\}^r$, we clearly obtain convex combinations of uniform averages over increasingly large ranges of $m_2 - m_1$ of the last expression above, and so appealing to the usual mean ergodic theorem for the \mathbb{Z}^r -action $S_{d+1} := T_1 \times T_2 \times \dots \times T_d$ in $L^2(\mu^{*d})$ we deduce that our above double averages converge to

$$\int_X \prod_{i=1}^d f_i \circ \pi_i \cdot \left(\lim_{M \rightarrow \infty} \frac{1}{M^r} \sum_{m \in \{1, 2, \dots, M\}^r} \left(\prod_{i=1}^d f_i \circ \pi_i \right) \circ S_{d+1}^m \right) d\mu^{*d}.$$

Setting

$$g := \lim_{M \rightarrow \infty} \frac{1}{M^r} \sum_{m \in \{1, 2, \dots, M\}^r} \left(\prod_{i=1}^d f_i \circ \pi_i \right) \circ S_{d+1}^m$$

this is precisely an integral of the form that we are assuming vanishes, as required. \square

We are now in a position to construct our pleasant extension.

Proof of Proposition 4.6 We need to find an extension $(\tilde{X}, \tilde{\Sigma}, \tilde{\mu}, \tilde{T})$ such that, setting

$$\Xi := \tilde{\Sigma}^{\tilde{T}_1} \vee \tilde{\Sigma}^{\tilde{T}_2=\tilde{T}_1} \vee \dots \vee \tilde{\Sigma}^{\tilde{T}_d=\tilde{T}_1},$$

we have

$$\frac{1}{|I_N|} \sum_{n \in a_N + I_N} (\tilde{f}_1 - \mathbb{E}_{\tilde{\mu}}[\tilde{f}_1 | \Xi]) \circ \tilde{T}_1^n \cdot \prod_{i=2}^d \tilde{f}_i \circ \tilde{T}_i^n \rightarrow 0 \quad \text{in } L^2(\tilde{\mu})$$

for any $\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_d \in L^\infty(\tilde{\mu})$. By Lemma 4.7, this will follow if we can guarantee instead that

$$\int_{\tilde{X}^d} \tilde{f}_1 \circ \tilde{\pi}_1 \cdot \left(\prod_{i=2}^d \tilde{f}_i \circ \tilde{\pi}_i \right) \cdot \tilde{g} \, d\tilde{\mu}^{*d} = \int_{\tilde{X}^d} \mathbb{E}_{\tilde{\mu}}[\tilde{f}_1 | \Xi] \circ \tilde{\pi}_1 \cdot \left(\prod_{i=2}^d \tilde{f}_i \circ \tilde{\pi}_i \right) \cdot \tilde{g} \, d\tilde{\mu}^{*d}$$

for every choice of $\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_d \in L^\infty(\tilde{\mu})$ and of another function $\tilde{g} \in L^\infty(\tilde{\mu}^{*d}|_{(\tilde{\Sigma}^{\otimes d})^{\tilde{S}_{d+1}}})$.

We shall show that this obtains for the inverse limit of a tower of extensions of (X, Σ, μ, T) constructed from the Furstenberg self-joinings themselves.

Step 1: construction of the extension Given the original system (X, Σ, μ, T) we define an extension $\psi^{(1)} : (X^{(1)}, \Sigma^{(1)}, \mu^{(1)}, T^{(1)}) \rightarrow (X, \Sigma, \mu, T)$ by setting $(X^{(1)}, \Sigma^{(1)}, \mu^{(1)}) := (X^d, \Sigma^{\otimes d}, \mu^{*d})$, $\psi^{(1)} := \pi_1$ and with the \mathbb{Z}^r -actions

$$\begin{aligned} T_1^{(1)} &:= S_{d+1}, \\ T_2^{(1)} &:= S_2, \\ &\vdots \\ T_d^{(1)} &:= S_d \end{aligned}$$

(note that we lift T_1 to S_{d+1} , rather than to S_1). We may now iterate this construction on the systems that emerge from it to build a whole tower of extensions $(X^{(m)}, \Sigma^{(m)}, \mu^{(m)}, T^{(m)}) \rightarrow (X^{(m-1)}, \Sigma^{(m-1)}, \mu^{(m-1)}, T^{(m-1)})$ for $m \geq 1$, where we set $(X^{(0)}, \Sigma^{(0)}, \mu^{(0)}, T^{(0)}) := (X, \Sigma, \mu, T)$. Note that since each $(X^{(m+1)}, \Sigma^{(m+1)}, \mu^{(m+1)})$ is the d -fold Furstenberg self-joining of $(X^{(m)}, \Sigma^{(m)}, \mu^{(m)})$, in addition to the factor map $\pi_1^{(m)}$ given by the projection onto the first coordinate in this self-joining it carries $d-1$ other such maps corresponding to the projections onto the other coordinates; let us denote these by $\psi_2^{(m)}, \psi_3^{(m)}, \dots, \psi_d^{(m)}$.

We will take $(\tilde{X}, \tilde{\Sigma}, \tilde{\mu}, \tilde{T})$ to be the inverse limit $\lim_{m \leftarrow} (X^{(m)}, \Sigma^{(m)}, \mu^{(m)}, T^{(m)})$, and show that this has the desired property. Write $\psi : \tilde{X} \rightarrow X$ for the overall factor map back onto the original probability space, $\theta_{(m)}^{(m')} : X^{(m')} \rightarrow X^{(m)}$ for the connecting projections of our inverse system, and also $\theta_{(m)} : \tilde{X} \rightarrow X^{(m)}$ for

the overall projection from the limit system, so that $\psi = \theta_{(0)}$. Write $\pi_i^{(m)}$ for the coordinate projections $(X^{(m)})^d \rightarrow X^{(m)}$ and $\tilde{\pi}_i$ for the coordinate projections $\tilde{X}^d \rightarrow \tilde{X}$. Finally, let

$$\Xi^{(m)} := (\Sigma^{(m)})^{T_1^{(m)}} \vee (\Sigma^{(m)})^{T_1^{(m)}=T_2^{(m)}} \vee \dots \vee (\Sigma^{(m)})^{T_1^{(m)}=T_d^{(m)}}$$

and

$$\Xi := \tilde{\Sigma}^{\tilde{T}_1} \vee \tilde{\Sigma}^{\tilde{T}_2=\tilde{T}_1} \vee \dots \vee \tilde{\Sigma}^{\tilde{T}_d=\tilde{T}_1},$$

combining Lemmas 2.1 and 2.2 we deduce that $\Xi = \bigvee_{m \geq 1} \theta_{(m)}^{-1}[\Xi^{(m)}]$.

We can depict the tower of systems constructed above in the following commutative diagram:

$$\begin{array}{ccc} (\tilde{X}, \tilde{\Sigma}, \tilde{\mu}) & \xleftarrow{\tilde{\pi}_1} & (\tilde{X}^d, \tilde{\Sigma}^{\otimes d}, \tilde{\mu}^{*d}) \\ \downarrow & & \downarrow \\ \vdots & & \vdots \\ \theta_{(2)}^{(3)} \downarrow & & (\theta_{(2)}^{(3)})^{\times d} \downarrow \\ (X^{(2)}, \Sigma^{(2)}, \mu^{(2)}) & \xleftarrow{\pi_1^{(2)}} & ((X^{(2)})^d, (\Sigma^{(2)})^{\otimes d}, (\mu^{(2)})^{*d}) \\ \downarrow & & \downarrow \\ \theta_{(1)}^{(2)} \downarrow & & (\theta_{(1)}^{(2)})^{\times d} \downarrow \\ (X^{(1)}, \Sigma^{(1)}, \mu^{(1)}) & \xleftarrow{\pi_1^{(1)}} & ((X^{(1)})^d, (\Sigma^{(1)})^{\otimes d}, (\mu^{(1)})^{*d}) \\ \downarrow & & \downarrow \\ \theta^{(1)} \downarrow & & (\theta^{(1)})^{\times d} \downarrow \\ (X, \Sigma, \mu) & \xleftarrow{\pi_1} & (X^d, \Sigma^{\otimes d}, \mu^{*d}) \end{array}$$

where, in addition, by construction we have

$$(X^{(m+1)}, \Sigma^{(m+1)}, \mu^{(m+1)}) = ((X^{(m)})^d, (\Sigma^{(m)})^{\otimes d}, (\mu^{(m)})^{*d})$$

for very $m \geq 0$ with the actions $T_i^{(m+1)}$ selected from among the $S_i^{(m)}$ as above, and under this identification the maps $\theta_{(m)}^{(m+1)}$ and $\pi_1^{(m)}$ agree. On the other hand, the maps $\pi_1^{(m+1)}$ and $(\theta_{(m)}^{(m+1)})^{\times d}$ do not agree.

Step 2: proof of pleasantness We will now prove that for any $\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_d \in L^\infty(\tilde{\mu})$ and $\tilde{g} \in L^\infty(\tilde{\mu}^{*d}|_{(\tilde{\Sigma}^{\otimes d})^{\tilde{S}_{d+1}}})$ we have

$$\int_{\tilde{X}^d} \tilde{f}_1 \circ \tilde{\pi}_1 \cdot \left(\prod_{i=2}^d \tilde{f}_i \circ \tilde{\pi}_i \right) \cdot \tilde{g} d\tilde{\mu}^{*d} = \int_{\tilde{X}^d} \mathbb{E}_{\tilde{\mu}}[\tilde{f}_1 | \Xi] \circ \tilde{\pi}_1 \cdot \left(\prod_{i=2}^d \tilde{f}_i \circ \tilde{\pi}_i \right) \cdot \tilde{g} d\tilde{\mu}^{*d}$$

By continuity in $L^2(\tilde{\mu})$ and the definition of inverse limit, we may assume further that there is some finite $m \geq 1$ such that $\tilde{f}_i = f_i \circ \theta_{(m)}$ and $\tilde{g} = g \circ \theta_{(m)}^{\times d}$ for some $f_1, f_2, \dots, f_d \in L^\infty(\mu^{(m)})$ and $g \in L^\infty((\mu^{(m)})^{*d}|_{(\Sigma^{(m)})^{S_{d+1}}})$. Given this the left-hand expression above can be re-written at level m as

$$\int_{(X^{(m)})^d} f_1 \circ \pi_1^{(m)} \cdot \left(\prod_{i=2}^d f_i \circ \pi_i^{(m)} \right) \cdot g d(\mu^{(m)})^{*d}.$$

For any $m' \geq m$, since $((X^{(m')})^d, (\Sigma^{(m')})^{\otimes d}, (\mu^{(m')})^{*d}) =: (X^{(m'+1)}, \Sigma^{(m'+1)}, \mu^{(m'+1)})$, the left-hand side above can also be re-written as

$$\begin{aligned} & \int_{\tilde{X}^d} (f_1 \circ \theta_{(m)} \circ \tilde{\pi}_1) \cdot \left(\prod_{i=2}^d f_i \circ \theta_{(m)} \circ \tilde{\pi}_i \right) \cdot (g \circ \theta_{(m)}^{\times d}) d\tilde{\mu}^{*d} \\ &= \int_{(X^{(m')})^d} ((f_1 \circ \theta_{(m)}^{(m')}) \circ \pi_1^{(m')}) \cdot \left(\prod_{i=2}^d (f_i \circ \theta_{(m)}^{(m')}) \circ \pi_i^{(m')} \right) \cdot (g \circ (\theta_{(m)}^{(m')})^{\times d}) d(\mu^{(m')})^{*d} \\ &= \int_{X^{(m'+1)}} ((f_1 \circ \theta_{(m)}^{(m')}) \circ \theta_{(m')}) \cdot \left(\prod_{i=2}^d (f_i \circ \theta_{(m)}^{(m')}) \circ \psi_i^{(m'+1)} \right) \cdot (g \circ (\theta_{(m)}^{(m')})^{\times d}) d\mu^{(m'+1)} \\ &= \int_{\tilde{X}} ((f_1 \circ \theta_{(m)}^{(m'+1)}) \circ \theta_{(m'+1)}) \\ & \quad \cdot \left(\prod_{i=2}^d (f_i \circ \theta_{(m)}^{(m')}) \circ \psi_i^{(m'+1)} \circ \theta_{(m'+1)} \right) \cdot (g \circ (\theta_{(m)}^{(m')})^{\times d} \circ \theta_{(m'+1)}) d\tilde{\mu}. \end{aligned}$$

Now, the function $(f_i \circ \theta_{(m)}^{(m')}) \circ \psi_i^{(m'+1)}$ is invariant under the \mathbb{Z}^r -action

$$T_i^{(m')}(T_1^{(m')})^{-1} \times T_i^{(m')}(T_2^{(m')})^{-1} \times \dots \times \text{id} \times \dots \times T_i^{(m')}(T_d^{(m')})^{-1} =: T_i^{(m'+1)}(T_1^{(m'+1)})^{-1}$$

for each $i = 2, 3, \dots, d$, and the function $g \circ (\theta_{(m)}^{(m')})^{\times d}$ is invariant under $S_{d+1}^{(m')} =: T_1^{(m'+1)}$, so in the last integral above all factors save the first are $\theta_{(m'+1)}^{-1}[\Xi^{(m'+1)}]$ -measurable, and so we may condition $f_1 \circ \theta_{(m)}^{(m'+1)}$ onto $\Xi^{(m'+1)}$ and conclude

overall that

$$\begin{aligned}
& \int_{\tilde{X}^d} (f_1 \circ \theta_{(m)} \circ \tilde{\pi}_1) \cdot \left(\prod_{i=2}^d f_i \circ \theta_{(m)} \circ \tilde{\pi}_i \right) \cdot g \circ \theta_{(m)}^{\times d} d\tilde{\mu}^{*d} \\
&= \int_{\tilde{X}} (\mathbb{E}[f_1 \circ \theta_{(m)}^{(m'+1)}] \mid \Xi^{(m'+1)}) \circ \theta_{(m'+1)} \\
&\quad \cdot \left(\prod_{i=2}^d (f_i \circ \theta_{(m)}^{(m')}) \circ \psi_i^{(m'+1)} \circ \theta_{(m'+1)} \right) \cdot (g \circ (\theta_{(m)}^{(m')})^{\times d} \circ \theta_{(m'+1)}) d\tilde{\mu}.
\end{aligned}$$

Since

$$\mathbb{E}[f_1 \circ \theta_{(m)}^{(m')} \mid \Xi^{(m')}] \circ \theta_{(m')} \rightarrow \mathbb{E}[f_1 \circ \theta_{(m)} \mid \Xi]$$

and hence

$$\mathbb{E}[f_1 \circ \theta_{(m)}^{(m'+1)} \mid \Xi^{(m'+1)}] \circ \theta_{(m'+1)} - \mathbb{E}[f_1 \circ \theta_{(m)}^{(m')} \mid \Xi^{(m')}] \circ \theta_{(m')} \rightarrow 0 \quad \text{in } L^2(\tilde{\mu}) \text{ as } m' \rightarrow \infty,$$

we next deduce that

$$\begin{aligned}
& \int_{\tilde{X}} (\mathbb{E}[f_1 \circ \theta_{(m)}^{(m'+1)}] \mid \Xi^{(m'+1)}) \circ \theta_{(m'+1)} \\
&\quad \cdot \left(\prod_{i=2}^d (f_i \circ \theta_{(m)}^{(m')}) \circ \psi_i^{(m'+1)} \circ \theta_{(m'+1)} \right) \cdot (g \circ (\theta_{(m)}^{(m')})^{\times d} \circ \theta_{(m'+1)}) d\tilde{\mu} \\
&- \int_{\tilde{X}} (\mathbb{E}[f_1 \circ \theta_{(m)}^{(m')}] \mid \Xi^{(m')}) \circ \theta_{(m')} \\
&\quad \cdot \left(\prod_{i=2}^d (f_i \circ \theta_{(m)}^{(m')}) \circ \psi_i^{(m'+1)} \circ \theta_{(m'+1)} \right) \cdot (g \circ (\theta_{(m)}^{(m')})^{\times d} \circ \theta_{(m'+1)}) d\tilde{\mu} \\
&\rightarrow 0
\end{aligned}$$

as $m' \rightarrow \infty$, and by the law of iterated conditional expectation this last expression is equal to

$$\begin{aligned}
& \int_{\tilde{X}} (\mathbb{E}[\mathbb{E}[f_1 \circ \theta_{(m)}^{(m')}] \mid \Xi^{(m')}] \circ \theta_{(m')}) \circ \theta_{(m'+1)} \\
&\quad \cdot \left(\prod_{i=2}^d (f_i \circ \theta_{(m)}^{(m')}) \circ \psi_i^{(m'+1)} \circ \theta_{(m'+1)} \right) \cdot (g \circ (\theta_{(m)}^{(m')})^{\times d} \circ \theta_{(m'+1)}) d\tilde{\mu}.
\end{aligned}$$

However, by exactly analogous reasoning to that above applied with m' in place of m and the collection of functions $\mathbb{E}[f_1 \circ \theta_{(m)}^{(m')} \mid \Xi^{(m')}] \circ \theta_{(m')}$, $f_i \circ \theta_{(m)} = (f_i \circ$

$\theta_{(m)}^{(m')} \circ \theta_{(m')}$ for $i = 2, 3, \dots, d$ and $g \circ \theta_{(m)}^{\times d} = (g \circ (\theta_{(m)}^{(m')})^{\times d}) \circ \theta_{(m')}^{\times d}$ we deduce that this is equal to

$$\begin{aligned} & \int_{\tilde{X}^d} (\mathbb{E}[f_1 \circ \theta_{(m)}^{(m')} \mid \Xi^{(m')}] \circ \theta_{(m')} \circ \tilde{\pi}_1) \cdot \left(\prod_{i=2}^d f_i \circ \theta_{(m)} \circ \tilde{\pi}_i \right) \cdot g \circ \theta_{(m)}^{\times d} d\tilde{\mu}^{*d} \\ & \rightarrow \int_{\tilde{X}^d} (\mathbb{E}[f_1 \circ \theta_{(m)} \mid \Xi] \circ \tilde{\pi}_1) \cdot \left(\prod_{i=2}^d f_i \circ \theta_{(m)} \circ \tilde{\pi}_i \right) \cdot g \circ \theta_{(m)}^{\times d} d\tilde{\mu}^{*d} \end{aligned}$$

as $m' \rightarrow \infty$, as required. \square

It is clear that the assertion of Theorem 1.1 must hold for any system if it holds for some extension of that system, and so, as remarked previously, it now follows in full generality by combining Proposition 4.5 and Proposition 4.6. \square

Remarks Intuitively, at each step in our iterative construction of the tower of extensions

$$(X, \Sigma, \mu, T) \leftarrow (X^{(1)}, \Sigma^{(1)}, \mu^{(1)}, T^{(1)}) \leftarrow (X^{(2)}, \Sigma^{(2)}, \mu^{(2)}, T^{(2)}) \leftarrow \dots$$

we are introducing a new supply of functions that are invariant under either $T_1^{(j)}$ or $T_1^{(j)}(T_i^{(j)})^{-1}$ that can contribute to building a conditional expectation of f_1 that will serve as a good approximation to it for the purpose of evaluating our integral. However, at each such step we introduce new functions on the larger system that we will also then need to handle in this way, and these will not be taken care of until the next extension. It is for this reason that the present construction relies on the passage all the way to an inverse limit.

Considering informally how the pleasant extension enables us to bring the proof of Proposition 4.5 to bear on a more general system, we can locate the concrete appearance of the extension $(\tilde{X}, \tilde{\Sigma}, \tilde{\mu}, \tilde{T})$ when we approximate f_1 by $\sum_{k=1}^K g_{1,k} \cdot g_{2,k} \cdots g_{d,k}$: the point is that while this sum overall approximates a function on the smaller system (X, Σ, μ, T) , the individual functions g_i that appear within it do not, and then when we separately replace composition with \tilde{T}_1^n by \tilde{T}_i^n for these functions this requires us to keep track of their individual orbits inside $L^\infty(\tilde{\mu})$, which will in general not be confined to $L^\infty(\mu)$. \triangleleft

5 Discussion

5.1 Alternative constructions of the extension

The scheme we have adopted to construct our pleasant inverse limit extension $(\tilde{X}, \tilde{\Sigma}, \tilde{\mu}, \tilde{T})$ of (X, Σ, μ, T) is far from canonical. In particular, there is more than one way to use some self-joining of (X, μ) built using the original transformations T to control the convergence of nonconventional averages, as we have done with the Furstenberg self-joining via Lemma 4.7. While this choice seems particularly well-adapted to giving a quick inductive proof of Theorem 1.1, it may be instructive to describe briefly an alternative such self-joining that could be used in a similar way. This is a simple generalization of the space $(X^{[d]}, \Sigma^{[d]}, \mu^{[d]})$ constructed by Host and Kra for their proof in [13] of Theorem 1.1 in the case of powers of a single transformation.

Given our original system (X, Σ, μ, T) , we construct a sequence of self-joinings $(X^{[1]}, \Sigma^{[1]}, \mu^{[1]}, T^{[1]})$, $(X^{[2]}, \Sigma^{[2]}, \mu^{[2]}, T^{[2]})$, \dots , $(X^{[d]}, \Sigma^{[d]}, \mu^{[d]}, T^{[d]})$, where each $(X^{[i]}, \Sigma^{[i]}, \mu^{[i]}, T^{[i]})$ is a 2^i -fold self-joining of (X, Σ, μ, T) , iteratively as follows. First set $(X^{[1]}, \Sigma^{[1]}) := (X^2, \Sigma^{\otimes 2})$ and let $\mu^{[1]}$ be the relatively independent self-joining $\mu \otimes_{\Sigma^{T_1}} \mu$ of μ over the isotropy factor Σ^{T_1} (see, for example, Section 6.1 of Glasner [10] for the general construction of relatively independent self-joinings). In addition, lift T_1 to $T_1 \times \text{id}_X$ and T_i to $T_i^{[1]} := T_i \times T_i$ for $i = 2, 3, \dots, d$. It is clear from our construction that these preserve $\mu^{[1]}$. Finally, let π_1 be the projection of X^2 onto the first coordinate. Now to form $(X^{[2]}, \Sigma^{[2]}, \mu^{[2]}, T^{[2]})$ we apply this construction to the system $(X^{[1]}, \Sigma^{[1]}, \mu^{[1]}, T^{[1]})$ but taking the relatively independent self-product of $\mu^{[1]}$ over the different isotropy factor $\Sigma^{T_1^{[1]}=T_2^{[1]}}$, and lifting $T_1^{[1]}$ to $T_1^{[1]} \times T_2^{[2]}$ and $T_i^{[1]}$ to $T_i^{[1]} \times T_i^{[1]}$ for $i = 2, 3, \dots, d$. We continue iterating this construction, at each step forming $(X^{[k]}, \Sigma^{[k]}, \mu^{[k]}, T^{[k]})$ by taking the relatively independent self-product over $\Sigma^{T_1^{[k-1]}=T_i^{[k-1]}}$ and lifting $T_1^{[k-1]}$ to $T_1^{[k-1]} \times T_k^{[k-1]}$ and $T_i^{[k-1]}$ to $T_i^{[k-1]} \times T_i^{[k-1]}$ for $i = 2, 3, \dots, d$, until we reach $k = d$. This gives the Host-Kra self-joining. Our convention is to index the 2^d -fold product $X^{[d]}$ that results by the power set $\mathcal{P}[d]$ (the set of all subsets of $\{1, 2, \dots, d\}$), so that $X^{[d]} = X^{\mathcal{P}[d]}$, in such a way that $X^{[1]}$ corresponds to the factor $X^{\{\emptyset, \{1\}\}}$ of this larger product, $X^{[2]}$ to the factor $X^{\{\emptyset, \{1\}, \{2\}, \{1, 2\}\}}$, and so on. In addition, we write $\pi_\alpha^{[d]}$ for the 2^d coordinate projections $X^{\mathcal{P}[d]} \rightarrow X$. We can now easily concatenate the above specifications to write out the resulting transformations $T_i^{[d]}$ in terms of

the original T_j : $T_1^{[d]} = \prod_{\alpha \in \mathcal{P}[d]} T_{1,\alpha}$ with

$$T_{1,\alpha} := \begin{cases} T_1 & \text{if } \alpha = \emptyset \\ \text{id}_X & \text{if } \alpha = \{1\} \\ T_i & \text{if } \max \alpha = i \text{ for } i = 2, 3, \dots, d, \end{cases}$$

and $T_i^{[d]}$ is simply $T_i^{\times \mathcal{P}[d]}$ for $i = 2, 3, \dots, d$.

This can serve as an alternative to the Furstenberg self-joining in light of the following lemma:

Lemma 5.1 (The Host-Kra self-joining controls nonconventional averages). *If $f_1 \in L^\infty(\mu)$ is such that*

$$\int_{X^{\mathcal{P}[d]}} f_1 \circ \pi_\emptyset \cdot \left(\prod_{\alpha \in \mathcal{P}[d] \setminus \{\emptyset\}} f_\alpha \circ \pi_\alpha \right) d\mu^{[d]} = 0$$

for every choice of $f_\alpha \in L^\infty(\mu)$ for $\alpha \in \mathcal{P}[d] \setminus \{\emptyset\}$, then also

$$\frac{1}{N} \sum_{n=1}^N \prod_{i=1}^d f_i \circ T_i^n \rightarrow 0$$

in $L^2(\mu)$ for every choice of $f_2, f_3, \dots, f_d \in L^\infty(\mu)$.

Proof This follows essentially by d times applying alternately the van der Corput estimate, just as in the proof of Lemma 4.7), and then the Cauchy-Schwarz inequality for the space $L^2(\mu)$. The argument is just as for the case of powers of a single transformation treated by Host and Kra in [13] (see their Theorem 12.1 and the construction of Section 4), and we omit the details. \square

Writing $(X^{(1)}, \Sigma^{(1)}, \mu^{(1)}, T^{(1)}) := (X^{[d]}, \Sigma^{[d]}, \mu^{[d]}, T^{[d]})$ we can now use the machinery of Host-Kra self-joinings to build a tower of extensions of (X, Σ, μ, T) and deduce that their inverse limit is pleasant, as we did using the Furstenberg self-joining in Proposition 4.6. This requires grouping together the various factors in the integrand of

$$\int_{X^{\mathcal{P}[d]}} f_1 \circ \pi_\emptyset \cdot \left(\prod_{\alpha \in \mathcal{P}[d] \setminus \{\emptyset\}} f_\alpha \circ \pi_\alpha \right) d\mu^{[d]} = 0$$

according to the partition $\mathcal{P}[d] \setminus \{\emptyset\} = \bigcup_{i=1}^d \{\alpha : \max \alpha = i\}$, noting that the above explicit description of $T_1^{[d]}$ tells us that $f_{\{1\}} \circ \pi_{\{1\}}^{[d]}$ is $T_1^{[d]}$ -invariant and that

$$\prod_{\alpha: \max \alpha = i} f_{\alpha} \circ \pi_{\alpha}^{[d]}$$

is $T_1^{[d]}(T_i^{[d]})^{-1}$ -invariant for $i = 2, 3, \dots, d$. The remaining details of the argument are almost identical to those for Proposition 4.6. We note that in this argument the one-step extension $(X^{(1)}, \Sigma^{(1)}, \mu^{(1)}, T^{(1)})$ is already the top member of a height- d tower of self-joinings. These two towers serve different purposes in the proof, and should not be confused: the d smaller extensions used to build up to $(X^{(1)}, \Sigma^{(1)}, \mu^{(1)}, T^{(1)})$ correspond to the d appeals to the van der Corput estimate during the proof of Lemma 5.1.

The choice between the Furstenberg and Host-Kra self-joinings certainly affects the structure of the pleasant extension that emerges, but seems to make little difference to the overall complexity of the proof, since we do not exploit any of this more particular structure. The advantage of the Host-Kra self-joining is that it does not require an iterative appeal to Theorem 1.1 for its proof, but on the other hand that is traded off into a more complicated, alternating appeal to the van der Corput estimate and the Cauchy-Schwarz inequality in the proof of Lemma 5.1, rather than the simple single application made to prove Lemma 4.7.

Looking beyond the above considerations, it may be interesting to search for a quicker way to pass directly to a pleasant extension:

Question Can we construct a pleasant extension $(\tilde{X}, \tilde{\Sigma}, \tilde{\mu}, \tilde{T})$ in a finite number of steps, without invoking an inverse limit? \triangleleft

Remark Since a preprint of this paper first appeared, Bernard Host has shown in [11] that by using the above Host-Kra self-joining, one iteration of the above construction suffices to produce a pleasant system: the passage to the inverse limit is already superfluous! His proof of this requires a slightly more delicate analysis than the work of our Subsection 4.2, but in fact it seems likely that it applies equally well to both self-joinings. \triangleleft

5.2 Possible further questions

During the course of proving Theorem 1.1 we have made essential use of the commutativity of \mathbb{Z}^r , in addition to the commutativity of the different actions T_1 ,

T_2, \dots, T_d . It is possible that our theorem could be generalized by considering the averages

$$\frac{1}{|I_N|} \sum_{\gamma \in a_N I_N} \prod_{i=1}^d f_i \circ T_i^\gamma$$

for d commuting actions T_1, T_2, \dots, T_d on (X, Σ, μ) of a more general amenable group Γ with a Følner sequence $(I_N)_{N \geq 1}$ and base-point sequence $(a_N)_{N \geq 1}$. In this case, if we mimic our straightforward construction of the Furstenberg self-joining, we obtain a measure μ^{*d} on X^d that is $T_1 \times T_2 \times \dots \times T_d$ -invariant, but it may not now be invariant under any of the diagonal actions $T_i^{\times d}$. It seems that that ideas of the present paper cannot yield this stronger result (if it is true at all) without some additional new insight.

Another generalization of Theorem 1.1 has been conjectured by Bergelson and Leibman in [2]:

Conjecture (Nilpotent nonconventional ergodic averages). *If $T : \Gamma \curvearrowright (X, \Sigma, \mu)$ is a probability-preserving action of a discrete nilpotent group Γ and $\gamma_1, \gamma_2, \dots, \gamma_d \in \Gamma$ then for any $f_1, f_2, \dots, f_d \in L^\infty(\mu)$ the nonconventional ergodic averages*

$$\frac{1}{N} \sum_{n=1}^N \prod_{i=1}^d f_i \circ T^{\gamma_i^n}$$

converge to some limit in $L^2(\mu)$.

I do not know whether the methods of the present paper can be brought to bear on this conjecture; it seems likely that considerable further new machinery would be needed here also.

In a different direction, it is unknown whether Theorem 1.1 holds with pointwise convergence in place of convergence in $L^2(\mu)$. The methods of the present paper seem to contribute very little to our understanding of this problem; crucially, while the Furstenberg self-joining allows us to prove that $f_1 - E_\mu[f_1 | \Xi]$ contributes negligibly to the $L^2(\mu)$ convergence of our averages inside the extended system, so that we can replace f_1 with $E_\mu[f_1 | \Xi]$, we currently know of no good way to control this approximation pointwise, as would be essential for any approach to the question of pointwise convergence using the machinery of pleasant extensions and their factors.

References

- [1] V. Bergelson. Weakly mixing PET. *Ergodic Theory Dynam. Systems*, 7(3):337–349, 1987.
- [2] V. Bergelson and A. Leibman. A nilpotent Roth theorem. *Invent. Math.*, 147(2):429–470, 2002.
- [3] V. Bergelson, R. McCutcheon, and Q. Zhang. A Roth theorem for amenable groups. *Amer. J. Math.*, 119(6):1173–1211, 1997.
- [4] J.-P. Conze and E. Lesigne. Théorèmes ergodiques pour des mesures diagonales. *Bull. Soc. Math. France*, 112(2):143–175, 1984.
- [5] J.-P. Conze and E. Lesigne. Sur un théorème ergodique pour des mesures diagonales. In *Probabilités*, volume 1987 of *Publ. Inst. Rech. Math. Rennes*, pages 1–31. Univ. Rennes I, Rennes, 1988.
- [6] J.-P. Conze and E. Lesigne. Sur un théorème ergodique pour des mesures diagonales. *C. R. Acad. Sci. Paris Sér. I Math.*, 306(12):491–493, 1988.
- [7] N. Frantzikinakis and B. Kra. Convergence of multiple ergodic averages for some commuting transformations. *Ergodic Theory Dynam. Systems*, 25(3):799–809, 2005.
- [8] H. Furstenberg. Ergodic behaviour of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. d’Analyse Math.*, 31:204–256, 1977.
- [9] H. Furstenberg and B. Weiss. A mean ergodic theorem for $\frac{1}{N} \sum_{n=1}^N f(T^n x)g(T^{n^2} x)$. In V. Bergelson, A. March, and J. Rosenblatt, editors, *Convergence in Ergodic Theory and Probability*, pages 193–227. De Gruyter, Berlin, 1996.
- [10] E. Glasner. *Ergodic Theory via Joinings*. American Mathematical Society, Providence, 2003.
- [11] B. Host. Ergodic seminorms for commuting transformations and applications. Preprint.
- [12] B. Host and B. Kra. Convergence of Conze-Lesigne averages. *Ergodic Theory Dynam. Systems*, 21(2):493–509, 2001.

- [13] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. Math.*, 161(1):397–488, 2005.
- [14] A. Leibman. Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold. *Ergodic Theory Dynam. Systems*, 25(1):201–213, 2005.
- [15] M. G. Nadkarni. *Spectral theory of dynamical systems*. Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]. Birkhäuser Verlag, Basel, 1998.
- [16] T. Tao. Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory and Dynamical Systems*, 28:657–688, 2008.
- [17] H. P. Towsner. Convergence of Diagonal Ergodic Averages. Preprint, available online at arXiv.org: 0711.1180, 2007.
- [18] Q. Zhang. On convergence of the averages $(1/N) \sum_{n=1}^N f_1(R^n x) f_2(S^n x) f_3(T^n x)$. *Monatsh. Math.*, 122(3):275–300, 1996.
- [19] T. Ziegler. A non-conventional ergodic theorem for a nilsystem. *Ergodic Theory Dynam. Systems*, 25(4):1357–1370, 2005.
- [20] T. Ziegler. Universal characteristic factors and Furstenberg averages. *J. Amer. Math. Soc.*, 20(1):53–97 (electronic), 2007.

DEPARTMENT OF MATHEMATICS
 UNIVERSITY OF CALIFORNIA, LOS ANGELES,
 LOS ANGELES, CA 90095-1555, USA

Email: timaustin@math.ucla.edu

Web: <http://www.math.ucla.edu/~timaustin>